



Kobayashi, H., Rohrbach, P. B., Scheichl, R., Wilding, N. B., & Jack, R. L. (2019). Correction of coarse-graining errors by a two-level method: Application to the Asakura-Oosawa model. *Journal of Chemical Physics*, 151(14), [144108 (2019)].  
<https://doi.org/10.1063/1.5120833>

Publisher's PDF, also known as Version of record

Link to published version (if available):  
[10.1063/1.5120833](https://doi.org/10.1063/1.5120833)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via AIP Publishing at <https://aip.scitation.org/doi/full/10.1063/1.5120833>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# Correction of coarse-graining errors by a two-level method: Application to the Asakura-Oosawa model <sup>EP</sup>

Cite as: J. Chem. Phys. **151**, 144108 (2019); <https://doi.org/10.1063/1.5120833>

Submitted: 24 July 2019 . Accepted: 29 August 2019 . Published Online: 08 October 2019

Hideki Kobayashi, Paul B. Rohrbach, Robert Scheichl, Nigel B. Wilding <sup>id</sup>, and Robert L. Jack <sup>id</sup>

## COLLECTIONS

<sup>EP</sup> This paper was selected as an Editor's Pick



View Online



Export Citation



CrossMark

## ARTICLES YOU MAY BE INTERESTED IN

Steering a solute between coexisting solvation states: Revisiting nonequilibrium work relations and the calculation of free energy differences

The Journal of Chemical Physics **151**, 144105 (2019); <https://doi.org/10.1063/1.5117780>

Compatible observable decompositions for coarse-grained representations of real molecular systems

The Journal of Chemical Physics **151**, 134115 (2019); <https://doi.org/10.1063/1.5116027>

Accuracy, precision, and efficiency of nonequilibrium alchemical methods for computing free energies of solvation. I. Bidirectional approaches

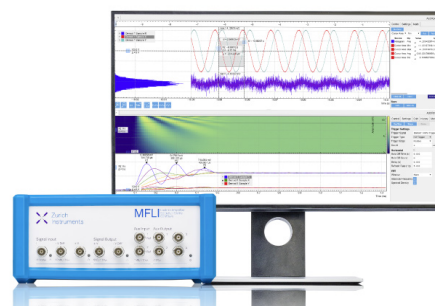
The Journal of Chemical Physics **151**, 144113 (2019); <https://doi.org/10.1063/1.5120615>

## Challenge us.

What are your needs for periodic signal detection?



Zurich  
Instruments



# Correction of coarse-graining errors by a two-level method: Application to the Asakura-Oosawa model

Cite as: J. Chem. Phys. 151, 144108 (2019); doi: 10.1063/1.5120833

Submitted: 24 July 2019 • Accepted: 29 August 2019 •

Published Online: 8 October 2019



Hideki Kobayashi,<sup>1</sup> Paul B. Rohrbach,<sup>2</sup> Robert Scheichl,<sup>3</sup> Nigel B. Wilding,<sup>4</sup>  and Robert L. Jack<sup>1,2,a)</sup> 

## AFFILIATIONS

<sup>1</sup>Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

<sup>2</sup>Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom

<sup>3</sup>Institute for Applied Mathematics, Heidelberg University, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany

<sup>4</sup>H.H. Wills Physics Laboratory, University of Bristol, Royal Fort, Bristol BS8 1TL, United Kingdom

<sup>a)</sup>Electronic mail: [rjl22@cam.ac.uk](mailto:rjl22@cam.ac.uk)

## ABSTRACT

We present a method that exploits self-consistent simulation of coarse-grained and fine-grained models in order to analyze properties of physical systems. The method uses the coarse-grained model to obtain a first estimate of the quantity of interest, before computing a correction by analyzing properties of the fine system. We illustrate the method by applying it to the Asakura-Oosawa model of colloid-polymer mixtures. We show that the liquid-vapor critical point in that system is affected by three-body interactions which are neglected in the corresponding coarse-grained model. We analyze the size of this effect and the nature of the three-body interactions. We also analyze the accuracy of the method as a function of the associated computational effort.

Published under license by AIP Publishing. <https://doi.org/10.1063/1.5120833>

## I. INTRODUCTION

Coarse-grained (CG) models are useful throughout many areas of physics and chemistry.<sup>1,2</sup> Depending on their context, they are used in different ways. In some cases, the idea is to develop a simple model that captures important qualitative features of some physical system.<sup>3–5</sup> Alternatively, one may develop a coarse-grained (CG) model that captures quantitatively the behavior of a more detailed (fine-grained, FG) model.<sup>2,6–9</sup> This second type of application is one of the foundations of multiscale modeling approaches.<sup>10,11</sup>

Enormous theoretical and computational effort has been invested in the development of accurate CG models. In the simplest case, one starts from a FG model with a classical Hamiltonian, and one aims to characterize its potential energy surface (PES) as a function of some simpler collective variables.<sup>6,12</sup> Alternatively, the FG model might be a quantum mechanical system, and one develops

a CG model that aims to fit the Born-Oppenheimer PES for the nuclear coordinates.<sup>10,11</sup> A number of recent studies have used machine-learning methods<sup>13–15</sup> to derive approximate PESs in this case.

Independent of the specific application, the most common approach in such cases is to derive a CG model that is as accurate as possible and then to estimate observable properties of the original (FG) model by computing the corresponding quantities in the CG model. If the CG model is an accurate representation of the fine one, then this approach works well. However, it is very rare for the CG model to be perfect, which means that this approach almost always results in some systematic errors, when estimating properties of the fine model. No information about these errors is available from analysis of the CG model alone.

In this work, we discuss a method for eliminating these coarse-graining errors. Specifically, we make quantitative estimates of

observable properties of the FG model, which we decompose as the value of the corresponding quantity for the CG model, and an explicit correction term. We show how this correction can be estimated in an efficient way. Our method is inspired by recent mathematical studies, particularly the multilevel approach,<sup>16–18</sup> which has been developed recently for applications in Bayesian statistics and uncertainty quantification.<sup>19–21</sup> (We note that the term “multilevel” has a specific mathematical meaning and should not be confused with the “multiscale” terminology used in physics and chemistry.<sup>10,11</sup>) Our method corresponds to a multilevel implementation with only two levels, so we call it a two-level (TL) method. It shares many features with importance sampling methods, biased sampling techniques,<sup>22,23</sup> and free-energy perturbation methods<sup>24</sup> (see, for example, Ref. 15 for a recent application of this method, in order to correct for coarse-graining errors). Our implementation of the TL method is informed by recent mathematical results and has been chosen to minimize the systematic errors (bias) in our estimates of properties of the FG system. As we discuss below, we minimize the bias by using a method for free-energy estimation<sup>25</sup> that is based on the fluctuation theorems of Jarzynski<sup>26</sup> and Crooks.<sup>27</sup> This free-energy estimation method is equivalent to the mathematical technique of annealed importance sampling.<sup>28</sup>

The method that we present is general. As a specific application, we use it to analyze the Asakura-Oosawa (AO) model<sup>29,30</sup> of colloid-polymer mixtures.<sup>31</sup> In this model, colloidal particles and polymers are described as spheres of different sizes, with pairwise interactions. The ratio of particle sizes is  $q \leq 1$ . In the corresponding CG model, only the colloidal particles are included. If  $q$  is sufficiently small then a CG model with pairwise interactions provides an exact description of the colloidal degrees of freedom.<sup>32</sup> For larger  $q$ , only approximate CG models are available because of effective interactions among the large particles that cannot be described by a pairwise additive potential.<sup>32–34</sup> In this case—as in many other CG models which restrict to pairwise interactions<sup>1,35</sup>—corrections to the CG model are required<sup>34</sup> in order to accurately estimate properties of the full (fine-grained) mixture.

For the specific case of the AO model, a method for calculation of these corrections was previously exploited by Dijkstra and co-workers.<sup>36,37</sup> They used a geometrical construction to analyze many-body interaction terms on the fly, during a CG simulation. Our method provides an alternative route to computation of these corrections: it provides an accurate characterization of the critical point of the AO mixture. As expected,<sup>33,36,37</sup> we find that many-body effective interactions for the colloids tend to reduce the tendency of the system to form clusters. We also analyze the nature and strength of the three-body and many-body effective interactions. In particular, for  $q = \frac{2}{5}$ , we show that a CG model with two- and three-body interactions can capture the critical point to high accuracy, where a two-body CG model fails.

The structure of the paper is as follows. Section II describes the general method, while Sec. III describes the AO model and details how the method is applied in this case. In Sec. IV, we present numerical results. Then, Sec. V analyses the accuracy of the method and its statistical uncertainties. Finally, Sec. VI summarizes the conclusions and outlook. Additional technical detail is provided in the appendices.

## II. THE TWO-LEVEL METHOD

### A. Outline

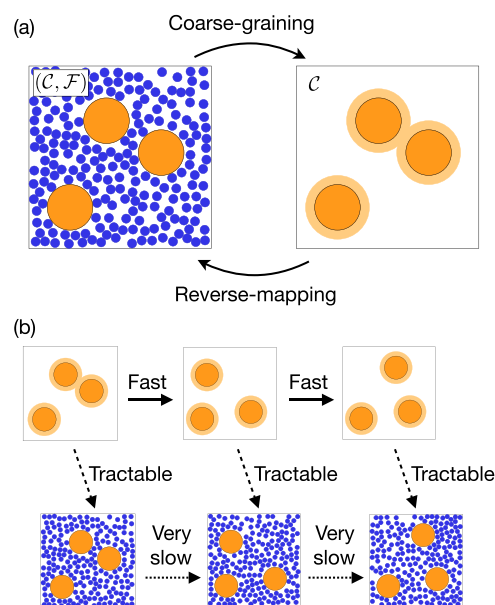
This section presents our method in general terms, using a schematic notation. The specific application that we consider is a mixture of large and small particles, for which full definitions are given in Sec. III.

Consider a system with many degrees of freedom and suppose that we are given a coarse-grained model which describes the same physical system in terms of a smaller set of (coarse) variables. In our application, the full system consists of a mixture of large and small particles, in the grand canonical ensemble. The coarse degrees of freedom are the number of large particles  $N$  and their positions  $(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N)$ . The other degrees of freedom are the number of small particles  $n$  and their positions  $(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n)$ . In our schematic notation, we denote a configuration of the coarse system as  $\mathcal{C} = (N, \mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N)$  and a configuration of the full system as  $(\mathcal{C}, \mathcal{F})$  with  $\mathcal{F} = (n, \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n)$ . This full system is the FG system, whose degrees of freedom comprise both  $\mathcal{C}$  and  $\mathcal{F}$ . See Fig. 1(a).

Given some configuration of this FG system, the energy is  $E(\mathcal{C}, \mathcal{F})$  and the associated Boltzmann distribution is

$$p_{\text{f}}(\mathcal{C}, \mathcal{F}) = \frac{1}{Z} e^{-\beta E(\mathcal{C}, \mathcal{F})}, \quad (1)$$

where  $\beta = 1/(k_{\text{B}}T)$  is the inverse temperature (scaled by Boltzmann's constant  $k_{\text{B}}$ ) and  $Z$  is the partition function.



**FIG. 1.** (a) Illustration of a FG mixture of large and small particles and a corresponding CG model that includes only the large particles, interacting by an effective potential. (b) TL method. Direct MC simulation of the FG model is very slow, so the TL method bypasses this step. It uses (fast) simulation of the CG model to generate independent configurations of the large particles. The small particles are then repopulated via a tractable step (similar to a reverse-mapping) that also yields an estimate for their free energy. This enables accurate computation of averages in the FG system.

The aim of the method is to compute averages with respect to  $p_f$ . We restrict to averages of observable quantities that depend only on the coarse degrees of freedom. The average of such a quantity  $A$  is denoted by

$$\langle A \rangle_f = \int p_f(C, \mathcal{F}) A(C) dC d\mathcal{F}. \quad (2)$$

The definition of the integration measure  $dC d\mathcal{F}$  depends on the system of interest. The marginal probability density for the coarse variables is

$$p_m(C) = \int p_f(C, \mathcal{F}) d\mathcal{F}. \quad (3)$$

Hence,  $\langle A \rangle_f = \int p_m(C) A(C) dC$ .

A configuration  $C$  of the coarse-grained system has energy  $E_c(C)$ . The Boltzmann distribution for the coarse system is

$$p_c(C) = \frac{1}{Z_c} e^{-\beta E_c(C)}, \quad (4)$$

and the associated equilibrium average is

$$\langle A \rangle_c = \int p_c(C) A(C) dC. \quad (5)$$

For a given fine system, the identification of suitable coarse degrees of freedom  $C$  and the inference of the energy function  $E_c(C)$  are both part of the coarse-graining procedure. A related procedure is the generation of representative configurations of the full (FG) system, starting from coarse configurations  $C$ ; this is known as reverse-mapping, see, for example, Ref. 38. See Fig. 1(a).

A great deal of effort<sup>1,2,8,9,13,14</sup> has gone into developing energy functions  $E_c$  such that coarse averages such as  $\langle A \rangle_c$  provide good estimates of the corresponding fine average  $\langle A \rangle_f$ . This obviously requires that  $p_c \approx p_m$ . Our method addresses the following question: Suppose that  $E$  and  $E_c$  are given, with  $p_c$  close (but not equal) to  $p_m$ . How should one obtain an accurate estimate of  $\langle A \rangle_f$  at reasonable computational cost?

The TL method can achieve this, as shown schematically in Fig. 1(b). It combines computations on the CG model with a reverse-mapping procedure that is used to assess the accuracy of this model. We first estimate  $\langle A \rangle_c$  by classical Monte Carlo (MC) sampling, which is assumed to have a relatively low computational cost. [In mathematics, this would be called a Markov chain Monte Carlo (MCMC) method.] This method is used to generate  $N_c$  representative samples from (4) which should be identically distributed but need not be independent. These are denoted by  $(C_1, C_2, \dots, C_{N_c})$ . Define the estimator

$$\hat{A}_c = \frac{1}{N_c} \sum_{i=1}^{N_c} A(C_i). \quad (6)$$

This is a consistent estimator of  $\langle A \rangle_c$ , but it is not an accurate estimate of the quantity of interest, which is  $\langle A \rangle_f$ . To quantify the difference between these two averages, we make a computational estimate of

$$\begin{aligned} \langle A \rangle_f - \langle A \rangle_c &= \int \left[ \frac{p_m(C)}{p_c(C)} - 1 \right] A(C) p_c(C) dC \\ &= \langle [w(C) - 1] A(C) \rangle_c, \end{aligned} \quad (7)$$

where  $w(C) = p_m(C)/p_c(C)$  is a reweighting factor. This is achieved by drawing  $N_f$  samples from those that were used in (1). With slight abuse of notation, we denote these by  $(C_1, C_2, \dots, C_{N_f})$ . Then, one estimates (7) as

$$\hat{\Delta} = \frac{1}{N_f} \sum_{i=1}^{N_f} [\hat{w}(C_i) - 1] A(C_i), \quad (8)$$

where  $\hat{w}(C_i)$  is an estimate of the ratio  $p_m(C_i)/p_c(C_i)$  that is obtained by a free energy computation using Jarzynski's equality (see Sec. III D; this step has significant computational cost but is tractable in the systems of interest). Finally, one has an estimator for the fine average

$$\hat{A} = \hat{A}_c + \hat{\Delta}. \quad (9)$$

We show below that as  $N_c, N_f \rightarrow \infty$ , then

$$\hat{A} \rightarrow \langle A \rangle_f. \quad (10)$$

## B. General properties

From (3), one identifies  $(-1/\beta) \log p_m(C)$  as the free energy of the  $\mathcal{F}$  variables, for a given coarse configuration  $C$ . Hence,  $E_c(C) = (-1/\beta) \log [Z_c p_c(C)]$  is an estimate of this free energy, and the correction (9) is analogous to the free-energy perturbation formula of Zwanzig.<sup>24</sup> In general, this perturbation formula suffers from large errors, except when the correction is small. Our TL method provides a method for computing the weights  $\hat{w}$  so as to obtain accurate estimates for  $\langle A \rangle_f$ .

It is natural to compare the TL method with simple MC algorithms that operate directly on the fine system. In cases where coarse-graining is useful, one should expect a separation of length scales between coarse and fine degrees of freedom. Often, this means that direct MC calculations on the fine system are frustrated by the disparate length scales. For example, it is likely that MC moves should involve coordinates that change only on the fine scale, but exploration of configuration space requires changes over the (much larger) coarse scale. The result is that it takes many such moves to relax (or decorrelate) the configurations of the system. In mathematical parlance, the MCMC chain for the FG system is expected to mix very slowly.

The TL method avoids this problem: the central idea is illustrated in Fig. 1(b). Note that the estimator  $\hat{A}_c$  only involves coarse degrees of freedom, so it does not suffer from the separation of length scales. Computation of  $\hat{\Delta}$  requires MC computations to be performed on the fine degrees of freedom, but it does not require that the coarse degrees of freedom are relaxed during this process. Hence, this computation does not suffer from the (very slow) mixing in the FG system. In cases where the CG model is accurate, this can lead to a strong improvement in performance.

Given a FG system, the TL method can be applied using any CG model. The results are (in principle) correct in all cases. This feature—which is standard within methods based on histogram reweighting, importance sampling, and free-energy perturbations<sup>23,24</sup>—allows enormous flexibility in the choice of the CG model. However, if the CG model is a poor representation of the fine one, the method is not efficient from a numerical point of view.



In this case, the weights  $\hat{w}_i$  acquire large fluctuations, which leads to large statistical errors on  $\hat{\Delta}$  (that is, large error bars). This is an in-built diagnostic of the quality of the CG model within the method. The computed weights can also be used to refine CG models to improve their accuracy.

### III. ASAKURA-OOSAWA MODEL SYSTEM

#### A. Fine-grained system

As an application of the TL method, we consider a model of a colloid-polymer mixture, as studied by Asakura and Oosawa<sup>29</sup> and by Vrij.<sup>39</sup> It consists of large and small spherical particles in three spatial dimensions, with diameters  $\sigma_B$  and  $\sigma_S$ , respectively. The behavior of this model is reviewed in Ref. 30; see also Fig. 2. Its phase behavior has been studied extensively, both computationally<sup>32,37</sup> and theoretically.<sup>40–44</sup> The large particles are hard spheres that represent the colloid. The small particles are schematic representations of (ideal) polymer chains, and their size is given by the polymer radius of gyration. These spheres are free to overlap with each other because the excluded volume of a polymer is very small in comparison with

its radius of gyration. In colloid-polymer mixtures, the effect of the polymers is to promote clustering of the colloidal particles, which leads eventually to phase separation and/or crystallization.<sup>31</sup> The AO model captures both these features.<sup>30</sup> We focus here on fluid phases.

The particles occupy a periodic box in three dimensions of size  $L \times L \times L$ . The large particles are hard: the potential energy  $U_f(\mathcal{C}, \mathcal{F}) = \infty$  if two large particles overlap or if a large particle overlaps with a small one; otherwise  $U_f(\mathcal{C}, \mathcal{F}) = 0$ . This means, in particular, that small particles are free to overlap with each other. The chemical potentials of the large and small particles are  $\mu_B$ ,  $\mu_S$ . The Boltzmann distribution for this system is similar to (1),

$$p_f(\mathcal{C}, \mathcal{F}) = \frac{1}{\Xi_f} e^{\beta\mu_B N + \beta\mu_S n - \beta U_f(\mathcal{C}, \mathcal{F})}, \quad (11)$$

where  $\Xi_f$  is the grand partition function. Computation of averages like (2) requires a definition of the integration measure  $d\mathcal{C}d\mathcal{F}$ : this is chosen such that the average in (2) is equal to an average over the grand canonical ensemble. It includes sums over  $N$ ,  $n$  and also a combinatorial factor of  $N!n!$  that accounts for particle indistinguishability. Details are given in Appendix A 1.

In place of the parameter  $\beta\mu_S$ , it is convenient to work with the volume fraction of the small particles  $\eta_S^r$  in a (nominal) reservoir at chemical potential  $\mu_S$ , that is,  $\eta_S^r = \pi e^{\beta\mu_S}/6$ . The size ratio between large and small particles is

$$q = (\sigma_S/\sigma_B). \quad (12)$$

Since the small particles do not interact with one another, the integral in (3) can be computed by summing over the number of small particles and integrating their positions. Following Ref. 32, we show in Appendix A 1 that for configurations  $\mathcal{C}$  without overlapping particles, the result is

$$p_m(\mathcal{C}) = \frac{1}{\Xi_f} \exp \left[ \beta\mu_B N + \frac{6\eta_S^r}{\pi\sigma_S^3} \mathcal{V}_a(\mathcal{C}) \right], \quad (13)$$

where  $\mathcal{V}_a(\mathcal{C})$  is the total volume that is accessible to the (centers of) the small particles.

We summarize the dimensionless model parameters, which are the colloid size ratio  $q$ , the large-particle chemical potential  $\beta\mu_B$ , the volume fraction of the small particles (in the reservoir)  $\eta_S^r$ , and the size of the simulation box  $L/\sigma_B$ .

#### B. Coarse-grained model

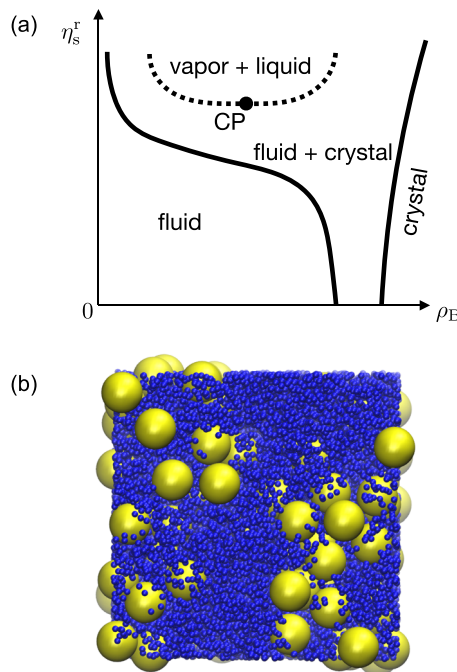
The coarse system only has large particles. These still have (repulsive) hard cores, but they also interact by a pairwise additive interaction with the Asakura-Oosawa (AO) pair potential. That is, for configurations  $\mathcal{C}$  without overlapping particles, one has<sup>30</sup>

$$\beta U_c(\mathcal{C}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \beta V_{AO}(|\mathbf{R}_i - \mathbf{R}_j|), \quad (14)$$

with

$$\beta V_{AO}(r) = \frac{-\eta_S^r}{q^3} \left[ (1+q)^3 - \frac{3r(1+q)^2}{2\sigma_B} + \frac{r^3}{2\sigma_B^3} \right]. \quad (15)$$

If there are overlapping particles, then  $U_c = \infty$ . The corresponding Boltzmann distribution is



**FIG. 2.** (a) Sketch of the phase diagram for the AO system,<sup>30,32</sup> as a function of the number density of large particles  $\rho_B$  and the reservoir volume fraction of small particles  $\eta_S^r$ . For  $q \lesssim 0.4$ , the liquid-vapor binodal occurs inside the fluid-crystal coexistence regime and is therefore metastable.<sup>32</sup> This work focuses on the (metastable) critical point marked by CP, which signals the onset of phase coexistence of the colloidal liquid and colloidal vapor phases. (b) Snapshot of the fine system at its critical point. The size ratio is  $q = (2/13) \approx 0.1538$  which is small enough that the coarse model is exact. The box size is  $L = 6\sigma_B$  and  $\eta_S^r = 0.3198$ , and the number of large particles in this configuration is  $N = 108$ . Direct simulation of this system is challenging due to the size disparity between large and small particles.

$$p_c(\mathcal{C}) = \frac{1}{\Xi_c} e^{\beta \mu_B N - \beta U_c(\mathcal{C})}, \quad (16)$$

where  $\Xi_c$  is the coarse partition function. Comparing (13) and (16), one sees that the coarse model corresponds to a pairwise additive approximation for the accessible volume  $\mathcal{V}_a$ . The model is parameterized by the same dimensionless parameters as the fine system ( $q, \beta \mu_B, \eta_S^f, L/\sigma_B$ ).

The behavior of this system depends on the size ratio  $q$ . In particular, for  $q \leq (2/\sqrt{3}) - 1 \approx 0.1547$ , a geometrical construction can be used to show that the pairwise-additive approximation for  $\mathcal{V}_a$  is exact and  $p_c(\mathcal{C}) = p_m(\mathcal{C})$  for all coarse configurations  $\mathcal{C}$ .<sup>30</sup> For larger values of  $q$ , the effective interaction in the fine system includes 3-body and higher contributions, which means that

$$\frac{p_m(\mathcal{C})}{p_c(\mathcal{C})} \propto \exp \left( \beta N \Delta \mu - \sum_{ijk} \beta V_3(\mathbf{R}_i, \mathbf{R}_j, \mathbf{R}_k) - \sum_{ijkl} \beta V_4(\mathbf{R}_i, \mathbf{R}_j, \mathbf{R}_k, \mathbf{R}_l) - \dots \right), \quad (17)$$

where the sums run over all the large particles,  $\Delta \mu$  is a constant, and  $V_3, V_4, \dots$  are many-body interaction potentials. For the specific case of the AO model, these many-body terms can be computed (numerically) using a geometrical construction, as shown in Refs. 36 and 37. For the purposes of this work, we note that  $p_m \neq p_c$  so computations on the coarse system (16) do not provide accurate estimates of properties of the fine system. Hence, the method of Sec. II is applicable.

The estimators (6) and (8) require that we generate representative samples from (16). This is achieved by grand canonical MC simulations of the coarse system.<sup>22</sup>

### C. Observable quantities

The phase diagram for the AO system is sketched in Fig. 2(a) as a function of  $\eta_S^f$  and the large particle number density  $\rho_B$  for a representative value of  $q$ . For small  $\eta_S^f$ , there is a fluid phase at low to moderate density and a crystalline phase at high density. On increasing  $\eta_S^f$  at moderate  $\rho_B$ , the crystalline phase becomes stable, but the fluid is still accessible as a metastable phase. On further increasing  $\eta_S^f$  within the metastable fluid, there is a critical point at  $(\eta_S^f)_c$ , and for  $\eta_S^f > (\eta_S^f)_c$ , one observes coexistence between phases with many large particles (“colloidal liquid”) and few large particles (“colloidal vapor”). This work concentrates on the behavior near the critical point. In this regime, the system depends very sensitively on interaction parameters, so it provides a challenging test case for the TL method. In our grand-canonical MC simulations of the coarse system, the fluid is sufficiently metastable that we do not observe crystallization (see also Appendix B).

In the following, we focus on the critical point. As a quantity of interest, we consider the histogram of the number of large particles. That is, we compute the probability to have exactly  $N_b$  large particles, which is

$$P_b(N_b) = \langle \delta_{N, N_b} \rangle_f. \quad (18)$$

This is of the required form (2). The distribution  $P_b$  has a characteristic form at the critical point.<sup>45</sup> In the vicinity of this point, it depends strongly on the model parameters. Hence, the accurate

determination of  $P_b$  is a stringent test of our method and allows us to locate the critical point of the fine system.

A previous study<sup>33</sup> has characterized this critical point by direct Monte Carlo computations on the fine system. However, despite the use of advanced MC methods,<sup>46</sup> that work was limited to relatively large  $q \geq \frac{2}{5}$ , otherwise the scale separation between the large and small particles makes such computations extremely expensive. Figure 2 shows a typical configuration of the FG system near the critical point for a small size ratio  $q = \frac{2}{13}$ . This picture illustrates that particle insertion MC moves for the large particles would be extremely challenging in this regime. Here, we are able to obtain accurate results for the fine system at  $q = \frac{1}{4}$ . This allows us to estimate effects of three-body interactions, which are not captured by the coarse model, although they could be analyzed following Ref. 37. Our method is also accurate at  $q = \frac{2}{13} \approx 0.15$ : this case is useful as a test of the method, but the coarse model is already exact at that state point, so it does not yield any new physical insight.

### D. Jarzynski integration and estimation of $P_b$

Computation of the weights  $\hat{w}$  in (8) involves grand canonical MC simulations where the large particles are held fixed, while small particles are inserted and removed. The system is divided into cells of size  $\sigma_S$  so that there are  $n_{\text{cell}}$  cells in total: this facilitates the computation of whether an attempted small-particle insertion is successful. We define a Monte Carlo sweep (MCS) as a sequence of  $n_{\text{cell}}$  attempted MC updates, where each update is chosen to be either a particle insertion or a particle removal, with equal probability.

The weight factors  $\hat{w}(C_i)$  are computed in two stages. We first calculate non-normalized weights  $\hat{W}(C_i)$  by a Monte Carlo (Jarzynski) integration method.<sup>25,28</sup> The procedure is described below where we also explain that for any fixed  $\mathcal{C}$ ,

$$\langle \hat{W}(\mathcal{C}) \rangle_I = \exp \left[ \frac{6\eta_S^f}{\pi\sigma_S^3} \mathcal{V}_a(\mathcal{C}) + \beta U_c(\mathcal{C}) \right]. \quad (19)$$

The average in this equation is an expectation value with respect to the Monte Carlo (Jarzynski) integration (see below). Then, we take

$$\hat{w}(C_i) = \frac{\hat{W}(C_i)}{\frac{1}{N_f} \sum_{j=1}^{N_f} \hat{W}(C_j)}. \quad (20)$$

Observe from (13) and (16) that (19) is equal to  $(\Xi_f/\Xi_c) \cdot (p_m(\mathcal{C})/p_c(\mathcal{C}))$ . Taking the average of (20) and comparing with (13) and (16), one has for large  $N_f$  that

$$\langle \hat{w}(\mathcal{C}) \rangle_I \approx \frac{p_m(\mathcal{C})}{p_c(\mathcal{C})}, \quad (21)$$

where we used that the denominator of (20) converges for large  $N_f$  to the expectation of (19) with respect to  $p_c$ , which is  $(\Xi_f/\Xi_c)$ . [The approximate equality in (21) is accurate up to a small correction associated with the difference between the denominator of (20) and its average; see Sec. V.]

We now describe the computation of the weights  $\hat{W}$ . Given some  $\mathcal{C}$ , we identify the grand partition function for the small particles as

$$\Xi_S(\mu_S, C) = \exp\left[e^{\beta\mu_S} \mathcal{V}_a(C)/\sigma_S^3\right] = \exp\left[\frac{6\eta_S^f}{\pi\sigma_S^3} \mathcal{V}_a(C)\right]. \quad (22)$$

To compute  $\hat{W}(C)$ , we hold the large particles fixed in position and we perform grand-canonical MC for the small particles. We first choose some chemical potential  $\mu_0$  which is small enough that the partition function  $\Xi_S(\mu_0, C)$  can be estimated directly from a grand canonical simulation of the small particles: it is equal to the reciprocal of the probability that  $n = 0$ .

Next, we perform an MC simulation with  $M$  MCS during which the small particle chemical potential increases from  $\mu_0$  to  $\mu_S$ . Let the small-particle chemical potential after  $t$  MCS be  $\mu(t)$ , and the number of small particles at this time be  $n(t)$ . Then, define

$$\mathcal{I}(C) = \int_0^M n(t) \frac{d}{dt} \mu(t) dt. \quad (23)$$

The parameter  $M$  and the dependence of  $\mu$  on  $t$  are arbitrary at this stage and can be chosen to optimize the accuracy of the method. In our simulations,  $\mu$  increases stepwise. Let  $\Delta\mu_j$  be the change in  $\mu$  on the  $j$ th step and  $n_j$  be the number of small particles in the system when that step takes place. Then,

$$\mathcal{I}(C) = \sum_{j=1}^K n_j \Delta\mu_j, \quad (24)$$

where  $K$  is the number of steps. (A simple choice is to make one step per MCS so that  $M = K$ , but other choices are possible.) A consistent choice of the weight is then  $\hat{W}(C) = \Xi_S(\mu_0, C) e^{\beta\mathcal{I}(C) + \beta U_c(C)}$ . Then, Crooks' fluctuation formula<sup>27</sup> (analogous to Jarzynski's equality<sup>26</sup>) can be used to show that (19) holds. A short derivation is given in Appendix A 2; see also Refs. 25, 28, and 47. Alternatively, we may repeat the same computation  $m$  times (always with the same  $C$ ) and take

$$\hat{W}(C) = \Xi_S(\mu_0, C) e^{\beta U_c(C)} \frac{1}{m} \sum_{p=1}^m e^{\beta \mathcal{I}_p(C)}, \quad (25)$$

where  $\mathcal{I}_p(C)$  is the value of (24) obtained on the  $p$ th iteration of the computation. Finally, using (25) in (20) gives the weight factors  $\hat{w}_i$ .

We summarize the parameters of the TL method. The numbers of configurations used for the coarse and fine estimates are  $\mathcal{N}_c, \mathcal{N}_f$ . These should both be large in order to have accurate results. In addition, the Jarzynski integration depends on  $M$  which is the total number of MCS in a Jarzynski run and  $m$  which is the number of Jarzynski runs for each  $C$ . The computation of  $\hat{\Delta}$  is the dominant computational cost in the method, which scales as  $mM\mathcal{N}_f$ . The choice of these parameters is discussed in Sec. V. The other choice that must be specified is the annealing schedule: that is, the dependence of  $\mu$  on  $t$  in (23), as well as the initial chemical potential  $\mu_0$  and the number of steps  $K$ . The choice of this schedule is discussed in Appendix B.

## IV. NUMERICAL RESULTS

### A. Critical behavior

As anticipated in Sec. III C, we focus on the probability distribution of  $N$ , the number of large particles in the system. We

analyze this distribution based on the theory of liquid-vapor critical points, but neglecting field-mixing effects.<sup>45</sup> At this level of theory, the distribution of  $N$  has a characteristic (universal) form at the critical point, with two peaks of equal height, separated by a trough. The ratio of probability between the peak and trough is 0.46, in contrast to the phase-coexistence regime for which this ratio scales with system size as  $e^{-\kappa L^2}$ , where  $\kappa$  is proportional to the surface tension. The two peaks correspond to particle densities whose difference scales as  $L^{-\beta/\nu}$ , where  $\beta, \nu$  are the critical exponents for the 3d Ising universality class, so that  $\beta/\nu \approx 0.52$ ; see, for example, Ref. 45. Hence, the variance of the distribution decreases as  $L^{-2\beta/\nu}$  as expected at liquid-vapor critical point. (These critical fluctuations are much larger than a comparable three-dimensional system with short-ranged correlations, where the variance decreases as  $L^{-3}$ .)

In order to characterize the shape of the distribution, we define a scaling variable  $X$  associated with the number of large particles by subtracting the mean of  $N$  and dividing by the standard deviation

$$\delta N = N - \langle N \rangle, \quad X = \frac{\delta N}{\langle \delta N^2 \rangle^{1/2}}. \quad (26)$$

The probability density for this variable is denoted by  $P_L(X)$ , where  $L$  indicates the system size.

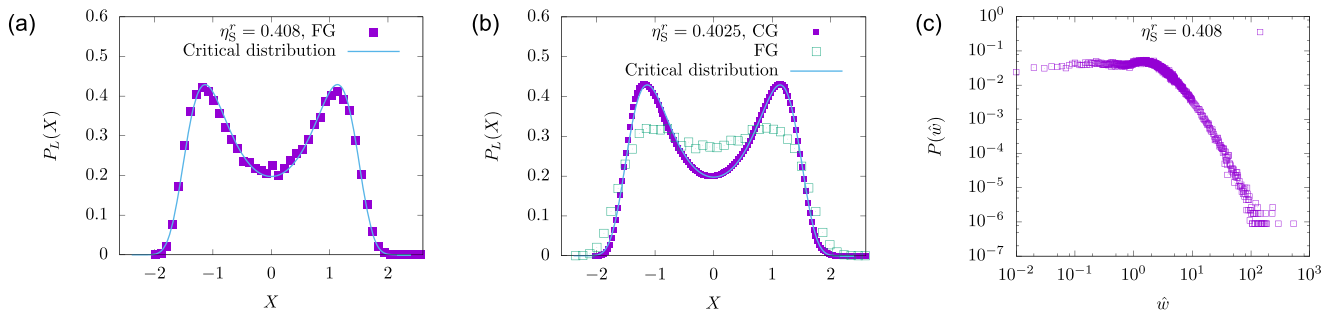
### B. Size ratio $q = \frac{1}{4}$

For  $q = (1/4)$ , direct simulation of the FG system is challenging, especially in the grand-canonical ensemble. However, the TL method is applicable. Figure 3 shows results obtained using this method, in a system of size  $L = 5\sigma_B$ . Figure 3(a) shows that the critical point in this system occurs at  $\eta_S^f \approx 0.4080$ . Taking this value and making an appropriate choice of  $\mu_B$ , the distribution of  $N$  matches well to the universal critical distribution.

Figure 3(b) shows results at  $\eta_S^f \approx 0.4025$  for both coarse and fine models. One sees that the coarse model is very close to its critical point, but the corresponding distribution for the fine model does not match the critical form. In the fine system, the bimodality of the distribution is less pronounced than the critical form, indicating that this system is in the one-phase regime, for this value of  $\eta_S^f$ . The physical interpretation of this result is that three-body (and higher) interactions in the fine system act to reduce the attractive forces, which pushes the critical point to a larger value of  $\eta_S^f$ . This shift is  $\Delta\eta_S^f \approx 0.006$  for these parameters. This is a small effect, but our method is accurate enough to resolve it. The qualitative difference between the FG distributions in Figs. 3(a) and 3(b) emphasizes that critical properties are very sensitive to model parameters: the values of  $\eta_S^f$  differ by less than 2%.

Note also that estimating the critical point by this method is only accurate up to an error that decays as a power law in  $L$ . This error has contributions from field-mixing effects and from corrections to scaling.<sup>45</sup> For the systems considered here, this correction might be comparable with  $\Delta\eta_S^f$  itself. However, we expect a similar correction for both coarse and fine systems, in which case our measured  $\Delta\eta_S^f$  should be a reasonable estimate of the difference in the critical points between coarse and fine systems.





**FIG. 3.** Results for  $q = (1/4)$  and  $L = 5\sigma_B$ . (a) Histogram of the large-particle density in the fine system (FG) at  $\eta_S^c = 0.4080$ , compared with the (universal) critical distribution. This result was obtained using the TL estimator  $\hat{A}$  defined in (9). (b) Histogram of the large-particle density in the coarse system (CG) at its critical point ( $\eta_S^c \approx 0.4025$ ), and the corresponding (reweighted) distribution for the fine system (FG) at the same value of  $\eta_S^c$ . From the shape of this distribution, one sees that the fine system is in the one-phase regime. (c) Histogram of the reweighting factors  $\hat{w}_i$  that are used in the estimator (8) when computing the results in panel (a).

For the system shown in Fig. 3(a), a histogram of the reweighting factors  $\hat{w}_i$  is shown in Fig. 3(c). If the coarse and fine models matched exactly and the Jarzynski integration was perfect, then all weight factors would be equal to unity. We see that some weights are significantly larger than 1: these are predominately coming from configurations whose probabilities are underestimated by the coarse model. Such configurations must be given larger weight when estimating observable quantities for the fine model, and this is the effect of the correction  $\hat{\Delta}$ .

### C. Size ratio $q = \frac{2}{5}$

We now turn to a size ratio  $q = (2/5)$ . This system was analyzed by direct simulation in Ref. 33, using the grand-canonical ensemble, with cluster moves<sup>46</sup> that remove small particles at the same time as inserting large ones (and vice versa). As in this work, the critical point was estimated by comparison with the universal order-parameter distribution: they found  $\eta_S^c = 0.5215 \pm 0.0001$  in a system of size  $L = 10\sigma_B$ . We note that the TL method does not outperform these previous methods for this moderate value of  $q$ : the TL method is designed to be effective for smaller  $q$ , which is the most challenging regime. For example, Sec. IV D considers  $q = (2/13)$ .

Reference 33 also shows that three-body interactions are significant for the size ratio  $q = (2/5)$ , in that the critical points differ significantly between the fine model and the coarse model with energy (14). We show in Appendix C that the differences between the coarse and fine systems are too large for the TL method to be able to correct the coarse-graining error in that case. We also explain how this problem can be avoided. Specifically, we define a new coarse model whose energy is

$$\beta U_{2,3}(C) = \beta U_c(C) + \beta U_3(C), \quad (27)$$

with  $U_c$  given by (14) and

$$U_3(C) = \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \sum_{k=j+1}^N V_3(R_{ij}, R_{ik}, R_{jk}), \quad (28)$$

where  $R_{ab} = |\mathbf{R}_a - \mathbf{R}_b|$  is the distance between particles  $a$  and  $b$ . That is,  $V_3$  is a 3-body interaction potential which we have chosen to

represent as a function of the three relevant interparticle distances [see Appendix C and recall also (17)].

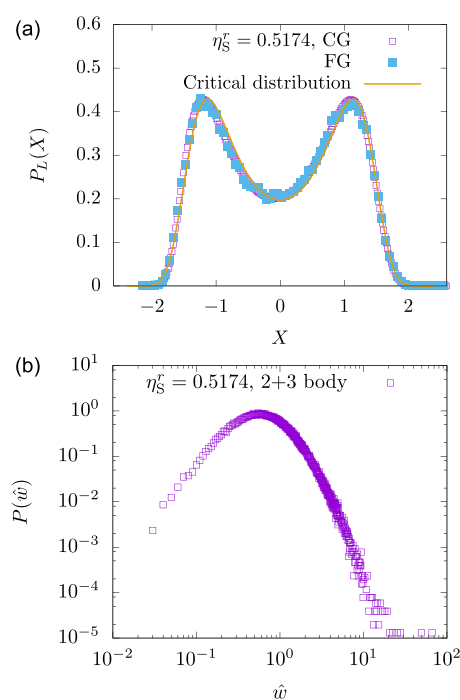
We refer to the model with energy (27) as the (2+3)-body coarse model. The three-body interaction is repulsive: it weakens the effective attractive attraction between particles, so the critical point for the (2+3)-body coarse model is at a higher volume fraction  $\eta_S^c$  than the critical point of the two-body coarse model (14).

Results using the (2+3)-body coarse model within the TL method are shown in Fig. 4. In particular, Fig. 4(a) shows histograms of the large-particle density for coarse and fine models at  $\eta_S^c = 0.5174$ , compared with the critical form. Two results are notable: first, the fine model matches well the critical distribution, so we conclude that this system is close to its critical point. Second, the fine model matches extremely accurately the coarse model. This indicates that the (2+3)-body coarse model is an almost-quantitative description of the fine system. This is confirmed by Fig. 4(b) which shows the distribution of  $\hat{w}_i$  which is peaked near  $\hat{w}_i = 1$ , consistent with the fact that the coarse-grained model is very accurate. Compared with Ref. 33, the difference (less than 1%) in the estimated critical value of  $\eta_S^c$  is presumably attributable to the smaller system size  $L = 6.4\sigma_B$  used in this work.

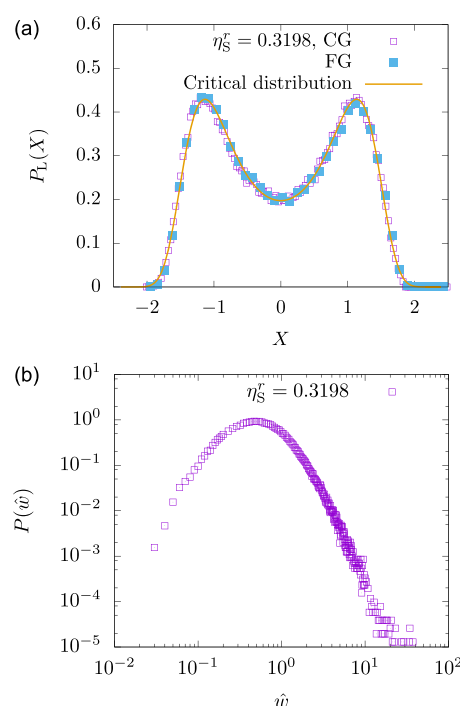
Since the (2+3)-body coarse model gives an almost-exact representation of the critical point, we conclude that four-body and higher interactions are negligible at this state point. This is perhaps not surprising since the (average) volume fraction of the large particles is around 0.22 at the critical point, and the effects of four-body and higher interactions are suppressed as the fourth power of the density. However, the result is promising in that it shows how low-order corrections to two-body interaction potentials can already give a significant improvement in accuracy.

### D. Size ratio $q = \frac{2}{13}$

As a third example system, we consider  $q = (2/13) \approx 0.1538$  for which the two-body coarse model is exact.<sup>30</sup> Results are shown in Fig. 5. They show that the FG and CG models do indeed have the same order-parameter distribution as they must. There are many



**FIG. 4.** Results for  $q = (2/5)$  and  $L = 6.4\sigma_B$ , using the TL method with the (2+3)-body coarse model. (a) Histogram of the large-particle density in the fine system (FG) at  $\eta_S^r = 0.5174$ , compared with the (universal) critical distribution and the corresponding distribution for the (2+3)-body coarse model (CG). All three distributions match very closely, indicating that the coarse model is an almost-exact representation of the fine model. (b) Histogram of reweighting factors. The distribution is peaked near  $\hat{w}_i = 1$ , which confirms that the (2+3)-body coarse model is an accurate description of the fine one.



**FIG. 5.** Results for  $q = (2/13)$  and  $L = 4.8\sigma_B$ . (a) Distributions of the number of large particles for coarse and fine models, compared with the universal critical distribution. (b) Histogram of weights. For this value of  $q$ , the coarse model is perfect so all of the variance of the weights comes from fluctuations in the Jarzynski integration.

small particles (recall Fig. 2, which has over 20 000 of them), so direct simulation of this system would be extremely challenging.

Within the TL method, it is known that  $\langle \hat{\Delta} \rangle = 0$  because the coarse model is exact. However, it is not easy to confirm this result numerically because the Jarzynski integration requires significant computational cost. Nevertheless, Fig. 5(b) shows that the distribution of weights  $\hat{w}_i$  is peaked near unity, which means that the method is effective even for this challenging test case. Figure 5(a) confirms that the results are accurate. The width of the distribution of weights comes entirely from the Jarzynski integration. It is suppressed by choosing a large value of  $M$ , corresponding to a significant computational effort for the computation of each individual weight. The small error on  $P(N)$  is ensured by combining a sufficiently narrow distribution of weights (large  $M$ ) with an average over many of these weights (sufficiently large  $\mathcal{N}_f$ ). For details of the simulations, see Appendix B. The scaling of the numerical error with respect to the algorithmic parameters is the subject of Sec. V.

The physical insight available from Fig. 5 is that the TL method is applicable even when the number of small particles is very large, recall Fig. 2(b).

## V. ERROR ESTIMATION AND CHOICE OF ALGORITHMIC PARAMETERS

### A. Overview

This section analyses the convergence shown in (10). We will show that as  $\mathcal{N}_c, \mathcal{N}_f \rightarrow \infty$ ,

$$\langle \hat{\Delta} - \langle A \rangle_f \rangle \rightarrow 0, \quad (29)$$

where the outer average (without subscript) is over all random aspects of the TL algorithm, which includes the generation of the original  $\mathcal{N}_c$  coarse samples, the selection of the  $\mathcal{N}_f$  samples to be used in (8), and the Jarzynski integration. The mean square error of the method is

$$h^2 = \langle (\hat{\Delta} - \langle A \rangle_f)^2 \rangle. \quad (30)$$

We will analyze the scaling of the error  $h^2$  in the limit  $\mathcal{N}_c, \mathcal{N}_f \rightarrow \infty$ .

Our analysis requires several assumptions, which are all satisfied in the physical situations of interest. Note first that the weighting factor  $w(C) = p_m(C)/p_c(C)$  may be very large if  $p_c$  is very small. This happens if there are configurations of the FG model that are very rare in the CG model. We assume that the variance of  $w(C_j)$  is finite, which requires that the CG model is not too inaccurate. In fact, it is useful to quantify the accuracy of the CG

model in terms of the  $\chi^2$ -divergence<sup>48,49</sup> between  $p_c$  and  $p_m$ , which is defined as

$$D_{\chi^2}(p_m \| p_c) = \int \left( \frac{p_m(C)}{p_c(C)} - 1 \right)^2 \cdot p_c(C) dC. \quad (31)$$

This quantity is non-negative, in general, and equals zero only if the coarse model is perfect, that is,  $p_c = p_m$ . Smaller values of  $D_{\chi^2}$  correspond to coarse models that are more accurate. Our assumption is that

$$D_{\chi^2}(p_m \| p_c) < \infty. \quad (32)$$

We also require an assumption on the variance of the weight factors  $\hat{W}$  generated by the Jarzynski integration method. As a minimal assumption, we suppose that this variance is finite, which means that

$$\langle \hat{W}(C_j)^2 \rangle < \infty. \quad (33)$$

The expected behavior of the weights  $\hat{W}$  is discussed below, but we expect that (33) holds in all situations of physical interest.

In addition to (32) and (33), we make several other assumptions to simplify the analysis. In particular, we assume that all the samples (both coarse and fine) are independent of each other. In practical situations, this assumption will not hold exactly, but we argue that our results still have the correct scaling with respect to the algorithmic parameters. For simplicity, we also assume throughout that  $A$  is a bounded quantity so that

$$|A(C) - \langle A \rangle_f| \leq a, \quad (34)$$

for some  $a$ , independent of  $C$ .

## B. Convergence of the mean

To establish (29), we first note from (6) that

$$\langle \hat{A}_c - \langle A \rangle_c \rangle = 0, \quad (35)$$

which follows by linearity of the average, because every  $C_i$  is distributed as  $p_c$ . Similarly,

$$\left\langle A(C_j) \frac{p_m(C_j)}{p_c(C_j)} \right\rangle = \langle A \rangle_f. \quad (36)$$

The next step is to quantify the approximation in (21). To this end, define  $\xi = \Xi_f / \Xi_c = \langle \hat{W}(C_j) \rangle$  and also

$$\epsilon = \frac{1}{\mathcal{N}_f} \sum_j [\hat{W}(C_j) / \xi] - 1. \quad (37)$$

Clearly,  $\langle \epsilon \rangle = 0$ , and  $\epsilon$  is an average of independent mean-zero variables. Also  $\hat{W}(C_j)$  has a finite variance by (33) so that  $\epsilon$  obeys a central limit theorem, and its typical value is therefore small as  $\mathcal{N}_f \rightarrow \infty$ . Also from (20),

$$\hat{w}(C_i) = \frac{\hat{W}(C_i)}{\xi} \left[ 1 - \frac{\epsilon}{1 + \epsilon} \right], \quad (38)$$

which is a more precise statement of (21). Observe that  $\sum_j \hat{w}(C_j) = 1$ , so (8) is identical to

$$\hat{\Delta} = \frac{1}{\mathcal{N}_f} \sum_j (A(C_j) - \langle A \rangle_f) (\hat{w}(C_j) - 1). \quad (39)$$

Hence, using again that  $C_j$  are distributed as  $p_c$ , one has

$$\langle \hat{\Delta} \rangle = \langle A \rangle_f - \langle A \rangle_c + \langle \hat{\epsilon} \rangle, \quad (40)$$

with

$$\hat{\epsilon} = \frac{1}{\mathcal{N}_f} \sum_j (A(C_j) - \langle A \rangle_f) \hat{w}(C_j). \quad (41)$$

Using (38) and that  $\epsilon$  is small, we can establish a bound on the average of  $\hat{\epsilon}$ . Details are given in Appendix D. The result is

$$\langle \hat{\Delta} \rangle = \langle A \rangle_f - \langle A \rangle_c + O(\mathcal{N}_f^{-1}). \quad (42)$$

With (9) and (35), this establishes (29).

We have emphasized throughout that the Jarzynski method for computing weights is useful because it minimizes systematic errors. To be specific, the simplest alternative to this approach would be to use a standard thermodynamic integration to estimate  $p_m/p_c$ , with the same computational effort  $M$  as was used in the Jarzynski method. The disadvantage of this approach is that one expects a systematic error on  $\hat{\Delta}$  from the finite rate of integration, which would appear in (42) as a correction at  $O(M^{-1})$ . (Recall that  $M$  is the total number of MCS used in the Jarzynski integration.) In this case, convergence of the method requires a joint limit  $\mathcal{N}_c, \mathcal{N}_f, M \rightarrow \infty$  instead of the (simpler) limit  $\mathcal{N}_c, \mathcal{N}_f \rightarrow \infty$  required here.

## C. Mean square error

This section estimates the error defined in (30), and how it scales with the parameters  $\mathcal{N}_c, \mathcal{N}_f, M$ . Let

$$h_c^2 = \langle (\hat{A}_c - \langle A \rangle_c)^2 \rangle. \quad (43)$$

Assuming that the  $\mathcal{N}_c$  samples used in  $\hat{A}_c$  are independent, it is immediate that

$$h_c^2 = \frac{\langle A^2 \rangle_c - \langle A \rangle_c^2}{\mathcal{N}_c}. \quad (44)$$

Using (34), the numerator is bounded above by  $a^2$ , so  $h_c^2 \rightarrow 0$  as  $\mathcal{N}_c \rightarrow \infty$ . In practice, (5) uses a sequence of samples from a MC simulation so that samples which are nearby in the sequence are not independent. However, these correlations will be short-ranged, as long as the MC simulation is converged. In this case,  $h_c^2$  is larger than (44) by a constant factor, but the scaling with  $\mathcal{N}_c$  is the same.

For the  $\mathcal{N}_f$  samples used in (8), it is again convenient to assume that these samples are independent of each other and of the coarse samples. The mean square error of  $\hat{\Delta}$  is

$$h_f^2 = \langle (\hat{\Delta} - \langle \hat{\Delta} \rangle + \langle \hat{\Delta} \rangle)^2 \rangle. \quad (45)$$

Writing  $\hat{\Delta} = (\hat{\Delta} - \langle \hat{\Delta} \rangle) + \langle \hat{\Delta} \rangle$  and using (42) yields

$$h_f^2 = \langle (\hat{\Delta} - \langle \hat{\Delta} \rangle)^2 \rangle + O(\mathcal{N}_f^{-2}). \quad (46)$$

We show below that  $h_f^2 = O(\mathcal{N}_f^{-1})$ , so it is consistent to drop the correction term. Using (39), we then obtain the leading order result

$$h_f^2 = \text{Var} \left( \frac{1}{\mathcal{N}_f} \sum_j (A(C_j) - \langle A \rangle_f) (\hat{w}(C_j) - 1) \right). \quad (47)$$

Substituting for  $\hat{w}$  with (38), the leading-order behavior can be obtained by dropping all terms proportional to  $\epsilon$ . Hence,

$$h_f^2 = \frac{1}{\mathcal{N}_f} \text{Var} [ (A(C_j) - \langle A \rangle_f) (\hat{W}(C_j)/\xi - 1) ]. \quad (48)$$

It follows from (34) that for any random variable  $B$ ,  $\text{Var}[(A(C_j) - \langle A \rangle_f)B(C_j)] \leq a^2 \langle B(C_j)^2 \rangle$ . Hence, (48) reduces to

$$h_f^2 \leq \frac{a^2}{\mathcal{N}_f} \left\langle \left( [\hat{W}(C_j)/\xi] - 1 \right)^2 \right\rangle. \quad (49)$$

The physical interpretation of (49) is that the scaling of the error is controlled by the variance of the weights  $\hat{W}$ . There are two contributing factors to this variance, which are the randomness in the Jarzynski integration and the fact that  $C_i$  is distributed as  $p_c$ .

We split the error estimate into two pieces, one from each source of randomness. Assume first that the reweighting factors  $\hat{w}$  are obtained exactly, so  $\hat{w}_i = p_m(C_i)/p_c(C_i)$ . In this case, all randomness comes from  $C_i$ . Hence, (49) reduces to

$$h_f^2 \leq h_{f,0}^2, \quad h_{f,0}^2 = \frac{a^2}{\mathcal{N}_f} D_{\chi^2}(p_m \| p_c). \quad (50)$$

The case where  $p_c = p_m$  is instructive, it corresponds to a perfect coarse model so that  $D_{\chi^2} = 0$  and  $h_{f,0}^2 = 0$ . Note however that the inequality in (50) is valid only if the weights  $\hat{w}$  are exact.

To account for the randomness in these weights, recall that the average of  $\hat{W}(C)$  obeys (19), but (48) also depends on the fluctuations of this quantity. To estimate the variance of  $\hat{W}(C)$ , note that  $\mathcal{I}$  in (24) is a sum of weakly correlated increments, so one expects its statistics to be ruled by a central limit theorem.<sup>28</sup> For this reason, we assume that for any fixed  $C$ , the sum  $\mathcal{I}$  has a Gaussian distribution with variance  $\Lambda/M$ . We assume that  $\Lambda$  is independent of  $C$  for convenience.

Recall from (25) that we obtain the weight  $\hat{W}$  by averaging over  $m$  independent realizations of the Jarzynski integration. If  $m = 1$ , one sees that  $\log \hat{W}(C)$  differs from  $\mathcal{I}$  by an additive constant and is therefore Gaussian with variance  $\Lambda/M$ . It follows that  $\langle \hat{W}(C)^2 \rangle_I = \langle W(C) \rangle_I^2 \exp(\Lambda/M)$ . For  $m > 1$ , the variance of  $\hat{W}$  is suppressed by a factor of  $m$ , which leads to

$$\langle \hat{W}(C)^2 \rangle_I - \langle \hat{W}(C) \rangle_I^2 = \frac{\langle \hat{W}(C) \rangle_I^2}{m} [\exp(\Lambda/M) - 1]. \quad (51)$$

Hence,

$$\left\langle \left( [\hat{W}(C)/\xi] - 1 \right)^2 \right\rangle_I = \left( \frac{p_m(C)}{p_c(C)} - 1 \right)^2 + \left( \frac{p_m(C)}{p_c(C)} \right)^2 \frac{\exp(\Lambda/M) - 1}{m}. \quad (52)$$

Using this result in (49) yields

$$h_f^2 \leq h_{f,0}^2 + h_{f,1}^2, \quad (53)$$

with

$$h_{f,1}^2 = \frac{a^2}{\mathcal{N}_f} [D_{\chi^2}(p_m \| p_c) + 1] \frac{\exp(\Lambda/M) - 1}{m}. \quad (54)$$

The key point here is that even if the coarse model is perfect ( $D_{\chi^2} = 0$ ), there is randomness in the Jarzynski integration which generates statistical errors. From (50), (53), and (54), one sees that indeed  $h_f^2 = O(\mathcal{N}_f^{-1})$ , as asserted above.

Combining all these ingredients and recalling our various assumptions, we find the scaling of the mean square error,

$$h^2 \leq a^2 \left[ \frac{1}{\mathcal{N}_c} + \frac{1}{\mathcal{N}_f} H_f \right], \quad (55)$$

with

$$H_f = D_{\chi^2}(p_m \| p_c) + [D_{\chi^2}(p_m \| p_c) + 1] \frac{\exp(\Lambda/M) - 1}{m}. \quad (56)$$

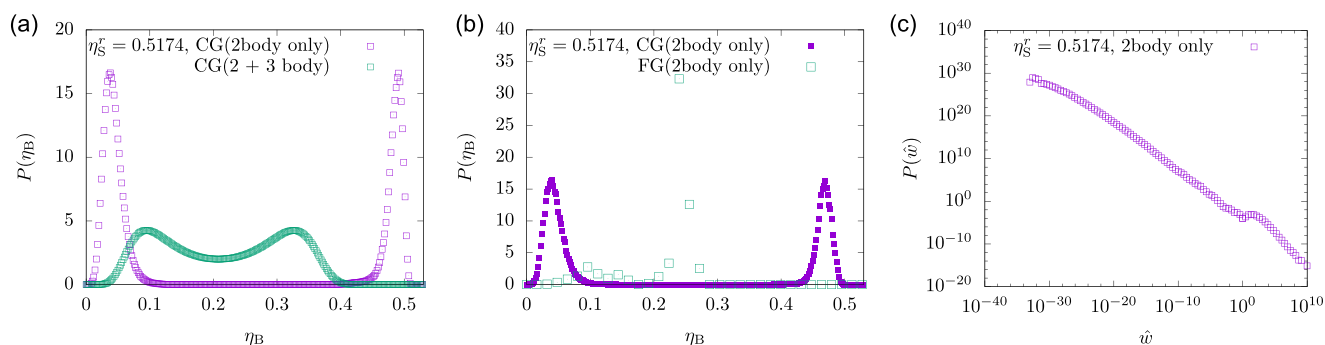
## D. Implications for optimization of the TL method

The results (55) and (56) establish useful properties of the TL algorithm. The computational effort for the Jarzynski integration is proportional to the product  $mM\mathcal{N}_f$ . We infer that the optimal parameter choice is to take  $\Lambda/M = O(1)$  in order that the exponential factor in (56) is not too large and then to devote the remaining computational effort toward larger  $\mathcal{N}_f$ . The benefits of increasing  $m$  are less than those of increasing  $\mathcal{N}_f$  or  $M$ , so the choice  $m = 1$  is likely optimal. However, it may be convenient in practice to take some other small integer since (for example) increasing  $m$  can be achieved by additional parallel processing, while increasing  $M$  requires longer simulation (wall) time. The parameters of the Jarzynski integration (particularly the annealing schedule) should be chosen to minimize  $\Lambda$  (at fixed  $M$ ). It is obviously desirable that  $D_{\chi^2}$  be small—this simply means that the coarse model should be as accurate as possible. In fact,  $D_{\chi^2}$  can be related to the exponential of a free energy difference between  $p_m$  and  $p_c$ , via (31). Hence, we expect that this quantity will scale exponentially in the system size—very accurate coarse models are required if the method is to be effective in very large systems.

The factors of  $D_{\chi^2}$  and  $e^{\Lambda/M}$  in (56) both indicate that the method can suffer from exponentially growing errors. However, the numerical results of Sec. IV show that the method is still practical in realistic soft-matter systems. The fact that  $D_{\chi^2}$  and  $\Lambda$  may be large in practice indicate that the majority of the computational effort in this method should be spent on the computation of  $\hat{\Delta}$ , which requires that both  $M$  and  $\mathcal{N}_f$  should be large. By contrast, we found that the error  $h_c$  coming from  $\mathcal{N}_c$  tends to be small, in the systems analyzed here.

## VI. CONCLUSION

We have presented a two-level method for correction of coarse-graining errors, and we have demonstrated it using the AO mixture. We have found that three-body interactions shift the critical point



**FIG. 6.** Results for  $q = (1/4)$  and  $L = 5\sigma_B$ , at  $\eta_S^r = 0.5174$  showing how the TL method breaks down when the coarse model is not sufficiently accurate. (a) Histogram of the large-particle density for the 2-body coarse model and the (2+3)-body coarse model. The coarse model with two-body interactions has two sharp peaks, showing phase coexistence. The FG system is close to criticality, for these parameters, as demonstrated in Fig. 4. The coarse model with (2+3)-body interactions is close to its critical point. (b) Estimate of the distribution for the number of large particles in the fine system obtained using (9), with the 2-body coarse model. In this case, the coarse model is not accurate and the estimated results for the fine system have large errors. (c) Corresponding distribution of reweighting factors. Most samples have extremely small weights, and a few have very large weights  $\hat{w}_i \gtrsim 10^8$ .

to higher  $\eta_S^r$  for  $q = (2/5)$  and  $q = (1/4)$ . For  $q = (2/5)$ , we characterized the three-body interactions, which have a significant effect on the critical point, and also we showed that four-body and higher interactions are negligible at criticality.

We also analyzed the errors associated with the method in Sec. V. The first important result of that section is that the method is accurate as  $\mathcal{N}_c, \mathcal{N}_f \rightarrow \infty$  but does not require any limit of large  $M$ . That is, we do not require that the individual weight factors  $\hat{w}_i$  are accurate measures of  $p_m/p_c$ . Instead, we require only that these weights have the right average, which is ensured by (21). The results (55) and (56) show how the parameters of the method should be chosen, as discussed in Sec. V D.

As noted above, the general two-level method is similar in spirit to free-energy perturbation theory and importance sampling. As such, it has a range of potential applications. Its main strength is that direct simulation of the whole system is never required. For example, in the system presented here, the large particles remain fixed, while the small particles are introduced, which avoids the difficulty of moving (or inserting) large particles in a system like Fig. 2. If one attempted direct MC simulation of the full system, then the error (55) would be inversely proportional to the total number of MCS, but it would be proportional to the number of MCS required to generate an independent sample of the full system. This number is expected to be extremely large in practice. (In particular, it would likely be much larger than  $\Lambda$ .) On the other hand, direct MC simulation does not suffer from problems associated with large values of  $D_{\chi^2}$ , which can hamper the TL method. That is, the TL method requires rather accurate coarse descriptions in order to be successful (see also Fig. 6 in Appendix C). Nevertheless, we have shown that information from the Jarzynski integration can be used to learn about corrections to coarse models, including three-body interactions. Combination of the TL method with machine-learning methods<sup>13–15</sup> seems to be a promising future direction.

Of course, there are many strategies for direct MC simulation. In particular, there are tailored algorithms for mixtures

of large and small particles, including the cluster moves used in Ref. 33, other cluster moves,<sup>50</sup> and staged (grand-canonical) particle insertion.<sup>51</sup> These methods reduce the number of MCS required to generate an independent sample. For the specific case of the AO model, the method of Ref. 37 is also available. A direct comparison with these methods is tricky because the computational efficiency depends strongly on details of the implementation. However, the results shown here are promising, especially given the relatively high volume fractions of the small particles. We also emphasize that the TL method is generic, while these other algorithms have been tailored to the specific problems of interest.

Another strength of the TL method is that while the computational effort of the Jarzynski integration is proportional to  $mM\mathcal{N}_f$ , this effort is easily split into  $m\mathcal{N}_f$  independent simulations, each involving  $M$  MCS. This means that the method is well-suited to parallel computation in contrast to direct MC simulation where very long individual simulations may be required.

In the future, we intend to test the TL method on mixtures of large and small hard spheres for which accurate coarse models can be computed theoretically.<sup>35,52,53</sup> The method will allow analysis of phase behavior in these systems and also quantification of the effects of many-body interactions.<sup>34</sup> Similar questions arise in systems where the large particles are not spherical, in which case the depletion interaction has a more complicated character<sup>54–56</sup> and can lead to new phase behavior.<sup>55,57</sup> There are also potential applications of the TL method to other soft matter systems including star polymers and dendrimers, for which accurate coarse-grained models are available,<sup>1,58,59</sup> and whose phase behavior is rich and complex.<sup>60,61</sup>

## ACKNOWLEDGMENTS

This work was supported by the Leverhulme Trust through research project Grant No. RPG–2017–203.



## APPENDIX A: ADDITIONAL THEORETICAL DETAILS

### 1. Grand canonical ensemble

To analyze these fluids in the grand canonical ensemble requires a definition of the integration measure in (2), which includes a sum over the numbers of particles and an integral over their positions. We take

$$\langle A \rangle_f = \frac{1}{\Xi_f} \sum_{N=0}^{\infty} \sum_{n=0}^{\infty} \frac{e^{\beta\mu_S n + \beta\mu_B N}}{n!N! \sigma_B^{3N} \sigma_S^{3n}} \times \int A(\mathcal{C}) e^{-\beta U_f(\mathcal{C}, \mathcal{F})} d\mathbf{R}_1 \dots d\mathbf{R}_N d\mathbf{r}_1 \dots d\mathbf{r}_n, \quad (\text{A1})$$

where all particle positions are integrated over box  $[0, L]^3$ . This formula is consistent with (11), recall that  $\mathcal{C} = (N, \mathbf{R}_1, \dots, \mathbf{R}_N)$  and  $\mathcal{F} = (n, \mathbf{r}_1, \dots, \mathbf{r}_n)$  and note that the zero of chemical potential has been chosen so that the thermal de Broglie wavelength does not appear, as usual in simulation studies. Note also that the definition of the integration measure  $d\mathcal{C}$  includes the factor of  $\frac{1}{N!}$  that accounts for indistinguishability of particles and similarly for  $d\mathcal{F}$ .

Also, normalization of  $p_f$  requires

$$\Xi_f = \sum_{N=0}^{\infty} \sum_{n=0}^{\infty} \frac{e^{\beta\mu_S n + \beta\mu_B N}}{n!N! \sigma_B^{3N} \sigma_S^{3n}} \int e^{-\beta U_f(\mathcal{C}, \mathcal{F})} d\mathbf{R}_1 \dots d\mathbf{R}_N d\mathbf{r}_1 \dots d\mathbf{r}_n. \quad (\text{A2})$$

Similar formulas hold for  $\langle A \rangle_c$  and  $\Xi_c$ .

To compute  $p_m(\mathcal{C})$  from (3) and (11), one notes that the integrations over the particle positions can be performed independently and that each of them yields a factor of  $\mathcal{V}_a$ , which is the volume accessible to a single small particle. For configurations  $\mathcal{C}$  where no large particles are overlapping, one finds

$$p_m(\mathcal{C}) = \frac{e^{\beta\mu_B N}}{\Xi_f} \sum_{n=0}^{\infty} \frac{e^{\beta\mu_S n}}{n! \sigma_S^{3n}} \mathcal{V}_a^n. \quad (\text{A3})$$

Recognizing the sum as the Taylor expansion of an exponential and using the definition of  $\eta_S^f$  yields (13).

### 2. Jarzynski average (19)

We outline the derivation of (19), following Ref. 27. As in (24), the chemical potential increases in a sequence of  $K$  steps. Hence, we consider a sequence of  $K + 1$  values of the chemical potential,  $(\mu_0, \mu_1, \dots, \mu_K)$ , with  $\mu_K = \mu_S$ , the value of the chemical potential for the fine system. Also,  $\Delta\mu_k = \mu_k - \mu_{k-1}$ . Between steps  $k$  and  $k + 1$ , one inserts and removes small particles using an MC algorithm that obeys detailed balance with respect to a probability distribution

$$p_k(\mathcal{C}, \mathcal{F}) \propto e^{\beta\mu_B N + \beta\mu_k n - \beta U_f(\mathcal{C}, \mathcal{F})}, \quad (\text{A4})$$

which is similar to (11) with  $\mu_S$  replaced by  $\mu_k$ . To establish (19) from (25), it is sufficient to show that

$$\langle e^{\beta \mathcal{I}(\mathcal{C})} \rangle_f = \frac{\Xi_S(\mu_S, \mathcal{C})}{\Xi_S(\mu_0, \mathcal{C})}, \quad (\text{A5})$$

where the average is with respect to this random MC algorithm and the small-particle partition function of (22) can be alternatively expressed as

$$\Xi_S(\mu, \mathcal{C}) = \int e^{\beta\mu n - \beta U_f(\mathcal{C}, \mathcal{F})} d\mathcal{F}. \quad (\text{A6})$$

Let the sequence of small-particle configurations when the steps take place be  $(\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_K)$ , and let the associated particle numbers be  $(n_1, n_2, \dots, n_K)$ . The initial small-particle configuration is distributed as

$$\pi_0(\mathcal{F}_1) = \frac{1}{\Xi_S(\mu_0, \mathcal{C})} e^{\beta\mu_0 n_1 - \beta U_f(\mathcal{C}, \mathcal{F}_1)}, \quad (\text{A7})$$

which is a normalized probability density. Then, write

$$\langle e^{\beta \mathcal{I}(\mathcal{C})} \rangle_f = \int \pi_0(\mathcal{F}_1) \left[ \prod_{k=1}^{K-1} q_k(\mathcal{F}_k \rightarrow \mathcal{F}_{k+1}) \right] \times e^{\sum_{k=1}^K \beta n_k (\mu_k - \mu_{k-1})} d\mathcal{F}_1 d\mathcal{F}_2 \dots d\mathcal{F}_K, \quad (\text{A8})$$

where  $q_k(\mathcal{F}_k \rightarrow \mathcal{F}_{k+1})$  is the probability (density) that the small particles evolve from  $\mathcal{F}_k$  to  $\mathcal{F}_{k+1}$ , between steps  $k$  and  $k + 1$ . Rearranging factors in the exponential term, this becomes

$$\langle e^{\beta \mathcal{I}(\mathcal{C})} \rangle_f = \int \left[ \prod_{k=1}^{K-1} q_k(\mathcal{F}_k \rightarrow \mathcal{F}_{k+1}) e^{\beta(n_k - n_{k+1})\mu_k} \right] \times \pi_0(\mathcal{F}_1) e^{-\beta(n_1 \mu_0 - n_K \mu_S)} d\mathcal{F}_1 d\mathcal{F}_2 \dots d\mathcal{F}_K. \quad (\text{A9})$$

Detailed balance of the Markov chains implies that

$$q_k(\mathcal{F}_k \rightarrow \mathcal{F}_{k+1}) e^{\beta(n_k - n_{k+1})\mu_k} = q_k(\mathcal{F}_{k+1} \rightarrow \mathcal{F}_k) \times \exp[-\beta U_f(\mathcal{C}, \mathcal{F}_{k+1}) + \beta U_f(\mathcal{C}, \mathcal{F}_k)]. \quad (\text{A10})$$

Using this in (A9) together with (A7) yields

$$\langle e^{\beta \mathcal{I}(\mathcal{C})} \rangle_f = \frac{1}{\Xi_S(\mu_0, \mathcal{C})} \int \left[ \prod_{k=1}^{K-1} q_k(\mathcal{F}_{k+1} \rightarrow \mathcal{F}_k) \right] \times e^{\beta\mu_S n_K - \beta U_f(\mathcal{C}, \mathcal{F}_K)} d\mathcal{F}_1 d\mathcal{F}_2 \dots d\mathcal{F}_K. \quad (\text{A11})$$

The integrals can now be performed one by one (starting with  $\mathcal{F}_1$ ) and using that  $q_k(\mathcal{F}_{k+1} \rightarrow \mathcal{F}_k)$  is a normalized probability density for  $\mathcal{F}_k$ . The final integral over  $\mathcal{F}_K$  is of the form (A6) with  $\mu = \mu_S$ , and one obtains (A5).

## APPENDIX B: SIMULATION DETAILS

For simulation of the coarse model, we use grand-canonical MC simulations as in Ref. 45. We restrict to particle insertion and removal (that is, no translational moves): this enables accurate results for fluid phases but helps to suppress crystallization.<sup>62</sup> We perform long MC runs (for example, up to  $10^8$  MCS, for which the computational effort is a few hours), and we extract  $\mathcal{N}_f = 128\,000$  configurations for the computation of  $\hat{\Delta}$ .

The Jarzynski integration follows Sec. III D and Appendix A 2. The parameter  $\mu_0$  is chosen such that

$$(L/\sigma_S)^3 e^{\beta\mu_0} = \bar{n}_0, \quad (\text{B1})$$

with  $\bar{n}_0$  of order unity being the average number of small particles that would be present, if there were no large particles. From grand-canonical simulations at this chemical potential (with fixed large particle configuration  $\mathcal{C}$ ), we estimate  $P_{\mathcal{C}}(n = 0)$  as the probability to have no small particles at all, from which we compute  $\Xi_S(\mu_0, \mathcal{C}) = 1/P_{\mathcal{C}}(n = 0)$ . The chemical potential is increased in steps with  $\beta\mu_k = \beta\mu_0 + \log(k + 1)$  so that  $(L/\sigma_S)^3 e^{\beta\mu_k} = (k + 1)\bar{n}_0$ . That is,

the average number of particles in the system increases linearly with the number of steps. Between each step in  $\mu$ , we perform 1 MCS for the small particles. We take  $\bar{n}_0 = (0.25, 0.48, 0.45)$  for  $q = (\frac{2}{5}, \frac{1}{4}, \frac{2}{13})$ . The number of steps  $K$  is fixed by the condition that  $\mu_K = \mu_S$ , which is the chemical potential for the state point of interest.

The nature of the AO system means that accurate computation of the Jarzynski integral can be simplified. The system is divided into cells of size  $\sigma_S$ . Before starting the computation, we identify cells that do not overlap at all with any large particle and also cells where the large particles block the insertion of any small particle. We do not attempt to insert small particles in any of these cells. The remaining cells are called *interaction cells*, and the number of them is  $n_{\text{int}}$ . Hence, the grand-canonical acceptance probabilities for particle insertion  $P_{\text{in}}$  and removal  $P_{\text{del}}$  are

$$P_{\text{in}} = \min\left(1, \frac{n_{\text{int}} v_{\text{cell}}}{(n' + 1) \sigma_S^3} \exp(\beta \mu_S)\right), \quad (\text{B2})$$

$$P_{\text{del}} = \min\left(1, \frac{n' \sigma_S^3}{n_{\text{int}} v_{\text{cell}}} \exp(-\beta \mu_S)\right), \quad (\text{B3})$$

where  $n'$  is the number of small particles in the system. The time taken for the Jarzynski integration depends strongly on the number of small particles that have to be inserted, but a single calculation might take a few minutes in a typical case.

We compute the Jarzynski integral based on (24) with this MC method, which yields an unbiased estimate  $e^{\beta \mathcal{I}(C)}$  of  $\exp[(e^{\beta \mu_S} - e^{\beta \mu_0}) \mathcal{V}_a^{\text{int}}(C) / \sigma_S^3]$ . Here,  $\mathcal{V}_a^{\text{int}}$  is the volume accessible to the small particles *within the interaction cells*. The total accessible volume is then  $\mathcal{V}_a = n_{\text{free}} v_{\text{cell}} + \mathcal{V}_a^{\text{int}}$ , where  $n_{\text{free}}$  is the number of cells that do not overlap at all with any large particle. This number is known, and so we can replace (25) with

$$\hat{W}(C) = \Xi_S(\mu_0, C) e^{\beta U_C(C)} \frac{1}{m} \sum_{p=1}^m e^{\beta \mathcal{I}_p(C)} \times \exp\left[(e^{\beta \mu_S} - e^{\beta \mu_0}) n_{\text{free}} v_{\text{cell}} / \sigma_S^3\right], \quad (\text{B4})$$

which still satisfies (19).

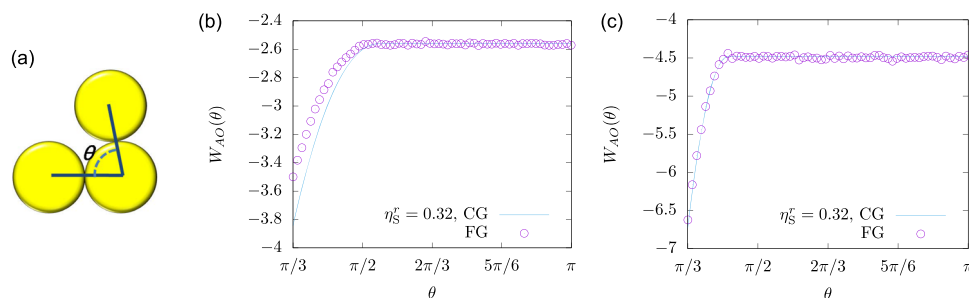
## APPENDIX C: COARSE MODEL WITH THREE-BODY INTERACTIONS

This Appendix gives extra information on three-body interactions in the system discussed in Sec. IV C, which has  $q = \frac{2}{5}$ . As a starting point, Fig. 6 shows the performance of the TL method, using the coarse model of (14), which has only two-body interactions. In contrast to Fig. 4, it is clear that the TL method is not able to correct the coarse-graining errors in this coarse model. The failure of the method appears as a small number of very large weights  $w_i$  in Fig. 6(c) up to  $10^{10}$ . The computational effort of Fig. 6 is comparable with that of Fig. 4, and  $N_f = 128\,000$  in both cases. In principle, the TL method can be used with any coarse model, but it is clear from (8) that  $N_f$  should be significantly larger than the largest weights appearing in the calculation, otherwise large errors are expected. Hence, one sees that the TL method is not feasible for the case shown in Fig. 6 since one would need  $N_f \gtrsim 10^{10}$ .

As discussed in Sec. IV C, we analyze the system at  $q = (2/5)$  using a CG model with (2+3)-body interactions. This interaction is represented as in (28), which is a scalar function  $U_3$  of three scalar variables (which are interparticle distances). The function is symmetric in all its arguments and is nonzero only if all arguments are between  $\sigma_B$  and  $\sigma_B + \sigma_S$ . In practice, we make the slightly stronger restriction that  $U_3(x, y, z) = 0$  except if  $\sigma_B^2 < (x^2 + y^2 + z^2) < (\sigma_B + \sigma_S)^2$ . We tabulate this function by estimating  $U_3(x_i, x_j, x_k)$  with  $x_j = j\Delta\sigma$  and  $\Delta\sigma = \sigma_S/10$ , and then we estimate the function at generic arguments using linear interpolation.

The tabulated values of the function are evaluated by the Jarzynski integration as in (19) for systems with  $N = 3$  large particles. We take  $L = 3.6\sigma_B$ , and we generate  $N_f = 512\,000$  configurations of the large particles by MC simulation of the coarse model (14). We perform the Jarzynski integration on each of these configurations  $C$  which yields an estimate of  $V_3(C) = 6\eta_S^3 \mathcal{V}_a(C) / (\pi\sigma_S^3\beta) + U_C(C)$  which is the three-body interaction in (17). For each  $C$ , we compute the distances  $R_{12}, R_{13}, R_{23}$  which are the arguments of  $U_3$ . Then, we define  $U_3(x_i, x_j, x_k)$  as an average the estimated  $V_3(C)$  over configurations  $C$  for which the 3-vector  $(R_{12}, R_{13}, R_{23})$  is inside a cubic box of side  $\Delta\sigma$ , centered at  $(x_i, x_j, x_k)$ .

An alternative approach would be to compute  $U_3$  based on a deterministically chosen set of large-particle configurations. This



**FIG. 7.** Three body interaction in the Asakura-Oosawa system as a function of  $\theta$  for  $\eta_S^c = 0.32$ . (a) Schematic illustration of the large-particle geometry used in these calculations. [(b) and (c)] The effective interaction in this three-particle system is shown with symbols (FG) in (a), calculated by the Jarzynski integration. The energy obtained from the pairwise additive two-body interaction potential of Eq. (14) is shown with a solid line (CG). In (b), we take  $q = \frac{1}{2}$  and there are significant differences between FG effective interaction and the two-body coarse model. In (c),  $q = \frac{1}{4}$  and the differences are smaller.

might reduce the computational effort, but it requires prior knowledge of the structure of the three-body potential. Instead, we computed the function at a set of representative points and we used these results to build the table. This “human-learning” method is sufficient for this simple three-body potential; machine-learning strategies would also be possible.<sup>13–15</sup>

To illustrate the structure of the three-body potential, Fig. 7 shows the energy of the (2+3)-body coarse model for three particles in contact as a function of the bond angle  $\theta$ . This calculation is performed by explicitly constructing large-particle configurations as in Fig. 7(c) and computing their weight factors using (19). Then,  $W_{AO}(C) = \frac{6\eta_s^f}{\pi\sigma_s^2} [\mathcal{V}_a(C) - \mathcal{V}_a(C_0)]$ , where  $C_0$  is a configuration where all particles in the system are well-separated. This is the exact coarse energy of the (2+3)-body system, which is to be compared with the energy of the two-body coarse model. Small  $\theta$  (close to  $\pi/3$ ) corresponds to the case where all three particles are close to each other. For  $q = (1/2)$ , the three-body interaction acts to increase the energy of such configurations, so the effective interaction between large particles is weaker than the prediction of the two-body coarse-model (14). For  $q = (1/4)$ , this effect is weaker although the result of Fig. 3(b) shows that even these small changes in the interaction potential have significant effects close to criticality.

#### APPENDIX D: THE BIAS OF THE ESTIMATOR $\hat{A}$

We show under the assumptions of Sec. V A that the expectation of  $\hat{\mathcal{E}}$  appearing in (40) is of order  $\mathcal{N}_f^{-1}$ . By definition,  $\hat{\mathcal{E}}$  is an importance sampling estimator which converges to 0 almost surely for  $\mathcal{N}_f \rightarrow \infty$ . In (38), we have seen that the normalization of  $\hat{w}$  slightly perturbs the weights appearing in  $\hat{\mathcal{E}}$ , i.e., in general,  $\epsilon \neq 0$ . This results in a bias which we can estimate by following Ref. 49. For completeness, we reproduce here the steps used in the proof of their Theorem 2.1, adapted to our notation.

We introduce the short-hand notation

$$\delta A(C_j) = A(C_j) - \langle A \rangle_f. \quad (D1)$$

Also note from (19) that

$$\langle \hat{W}(C_i)/\xi \rangle = \frac{p_m(C_i)}{p_c(C_i)}. \quad (D2)$$

Using the definition of  $\hat{\mathcal{E}}$  with (36), (38), and (D2) yields

$$\langle \hat{\mathcal{E}} \rangle = \left\langle \frac{-\epsilon}{1+\epsilon} \frac{1}{\mathcal{N}_f} \sum_j \delta A(C_j) \hat{W}(C_j)/\xi \right\rangle. \quad (D3)$$

We decompose the expectation value into contributions from  $\epsilon \leq -\frac{1}{2}$  and  $\epsilon > -\frac{1}{2}$  so that

$$|\langle \hat{\mathcal{E}} \rangle| \leq |\langle \hat{\mathcal{E}} 1_{\{2(\epsilon+1)>1\}} \rangle| + |\langle \hat{\mathcal{E}} 1_{\{2(\epsilon+1)\leq 1\}} \rangle|, \quad (D4)$$

where the indicator function  $1_{\{X\}}$  is equal to unity if  $X$  is true, and zero otherwise. For the first term on the right-hand side of (D4), we use that  $|\epsilon/(1+\epsilon)| \leq 2|\epsilon|$  for  $\epsilon > -\frac{1}{2}$  to obtain

$$|\langle \hat{\mathcal{E}} 1_{\{2(\epsilon+1)>1\}} \rangle| \leq 2 \left\langle |\epsilon| \cdot \left| \frac{1}{\mathcal{N}_f} \sum_j \delta A(C_j) \hat{W}(C_j)/\xi \right| \right\rangle. \quad (D5)$$

The right-hand side is less than or equal to

$$2\langle \epsilon^2 \rangle^{1/2} \cdot \left\langle \left( \frac{1}{\mathcal{N}_f} \sum_j \delta A(C_j) \hat{W}(C_j)/\xi \right)^2 \right\rangle^{1/2}, \quad (D6)$$

which follows from the Cauchy-Schwarz inequality. Using this fact together with (37) and that the  $C_i$  are all independent, we find

$$|\langle \hat{\mathcal{E}} 1_{\{2(\epsilon+1)>1\}} \rangle| \leq \frac{2}{\mathcal{N}_f \xi^2} \left( \langle \hat{W}(C_j)^2 \rangle \cdot \langle [\delta A(C_j) \hat{W}(C_j)]^2 \rangle \right)^{1/2}. \quad (D7)$$

Then, (34) yields

$$|\langle \hat{\mathcal{E}} 1_{\{2(\epsilon+1)>1\}} \rangle| \leq \frac{2a}{\mathcal{N}_f \xi^2} \langle \hat{W}(C_j)^2 \rangle. \quad (D8)$$

For the second term on the right-hand side of (D4), one uses (20), (34), and (41) to show that  $|\hat{\mathcal{E}}| \leq a$ . Hence,

$$|\langle \hat{\mathcal{E}} 1_{\{2(\epsilon+1)\leq 1\}} \rangle| \leq a \text{Prob}(\epsilon \leq -1/2). \quad (D9)$$

This probability can be bounded by Chebyshev's inequality which implies that for any mean-zero random variable  $X$  and any number  $x > 0$ ,  $\text{Prob}(X \leq -x) \leq \text{Var}(X)/x^2$ . Hence,

$$|\langle \hat{\mathcal{E}} 1_{\{2(\epsilon+1)\leq 1\}} \rangle| \leq 4a \text{Var}(\epsilon) = \frac{4a}{\xi^2 \mathcal{N}_f} \text{Var}(\hat{W}(C_j)). \quad (D10)$$

Combining (D4), (D8), and (D10), we have

$$|\langle \hat{\mathcal{E}} \rangle| \leq \frac{6a}{\xi^2 \mathcal{N}_f} \langle \hat{W}(C_j)^2 \rangle, \quad (D11)$$

which is an explicit bound that holds for all  $\mathcal{N}_f$ . The right-hand side is finite by (33) so  $\langle \hat{\mathcal{E}} \rangle$  is  $O(\mathcal{N}_f^{-1})$  as advertised above. With (40), this establishes (42).

We note in passing that if the weights have zero variance, then  $\hat{\mathcal{E}} = 0$  strictly and there is no bias although the bound (D11) does not capture this fact. In fact, this bound may be improved by refining the estimate of  $\langle \epsilon^2 \rangle$  used in (D7).

#### REFERENCES

- C. N. Likos, “Effective interactions in soft condensed matter physics,” *Phys. Rep.* **348**, 267 (2001).
- A. J. Pak and G. A. Voth, “Advances in coarse-grained modeling of macromolecular complexes,” *Curr. Opin. Struct. Biol.* **52**, 119 (2018).
- M. Hagan and D. Chandler, “Dynamic pathways for viral capsid assembly,” *Biophys. J.* **91**, 42 (2006).
- T. K. Haxton and S. Whitlam, “Design rules for the self-assembly of a protein crystal,” *Soft Matter* **8**, 3558 (2012).
- A. Saric, Y. C. Chebaro, T. P. J. Knowles, and D. Frenkel, “Crucial role of nonspecific interactions in amyloid nucleation,” *Proc. Natl. Acad. Sci. U. S. A.* **111**, 17869 (2014).
- W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, “The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models,” *J. Chem. Phys.* **128**, 244114 (2008).
- M. Praprotnik, L. Delle Site, and K. Kremer, “A macromolecule in a solvent: Adaptive resolution molecular dynamics simulation,” *J. Chem. Phys.* **126**, 134902 (2007).
- T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, “Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model,” *J. Chem. Phys.* **134**, 085101 (2011).
- B. M. Mladek, J. Fornleitner, F. J. Martinez-Veracoechea, A. Dawid, and D. Frenkel, “Procedure to construct a multi-scale coarse-grained model of DNA-coated colloids from experimental data,” *Soft Matter* **9**, 7342 (2013).

- <sup>10</sup>M. Karplus, "Development of multiscale models for complex chemical systems: From H+H<sub>2</sub> to biomolecules (Nobel lecture)," *Angew. Chem., Int. Ed.* **53**, 9992 (2014).
- <sup>11</sup>A. Warshel, "Multiscale modeling of biological functions: From enzymes to molecular machines (Nobel lecture)," *Angew. Chem., Int. Ed.* **53**, 10020 (2014).
- <sup>12</sup>S. T. John and G. Csányi, "Many-body coarse-grained interactions using Gaussian approximation potentials," *J. Phys. Chem. B* **121**, 10934 (2017).
- <sup>13</sup>J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," *Phys. Rev. Lett.* **98**, 146401 (2007).
- <sup>14</sup>A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, "Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons," *Phys. Rev. Lett.* **104**, 136403 (2010).
- <sup>15</sup>B. Cheng, E. A. Engel, J. Behler, C. Dellago, and M. Ceriotti, "Ab initio thermodynamics of liquid and solid water," *Proc. Natl. Acad. Sci. U. S. A.* **116**, 1110 (2019).
- <sup>16</sup>S. Heinrich, "Monte Carlo complexity of global solution of integral equations," *J. Complexity* **14**, 151 (1998).
- <sup>17</sup>M. B. Giles, "Multi-level Monte Carlo path simulation," *Oper. Res.* **56**, 607 (2008).
- <sup>18</sup>D. F. Anderson and D. J. Higham, "Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics," *Multiscale Model. Simul.* **10**, 146 (2012).
- <sup>19</sup>V. H. Hoang, C. Schwab, and A. M. Stuart, "Complexity analysis of accelerated MCMC methods for Bayesian inversion," *Inverse Probl.* **29**, 85010 (2013).
- <sup>20</sup>T. J. Dodwell, C. Ketelsen, R. Scheichl, and A. L. Teckentrup, "A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow," *SIAM/ASA J. Uncertainty Quantif.* **3**, 1075 (2015).
- <sup>21</sup>A. Beskos, A. Jasra, K. Law, R. Tempone, and Y. Zhou, "Multilevel sequential Monte Carlo samplers," *Stochastic Processes Their Appl.* **127**, 1417 (2017).
- <sup>22</sup>D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Elsevier, 2001).
- <sup>23</sup>A. D. Bruce and N. B. Wilding, "Computational strategies for mapping equilibrium phase diagrams," *Adv. Chem. Phys.* **127**, 1 (2003).
- <sup>24</sup>R. W. Zwanzig, "High-temperature equation of state by a perturbation method. I. Nonpolar gases," *J. Chem. Phys.* **22**, 1420 (1954).
- <sup>25</sup>G. Hummer and A. Szabo, "Free energy reconstruction from nonequilibrium single-molecule pulling experiments," *Proc. Natl. Acad. Sci. U. S. A.* **98**, 3658 (2001).
- <sup>26</sup>C. Jarzynski, "Nonequilibrium equality for free energy differences," *Phys. Rev. Lett.* **78**, 2690 (1997).
- <sup>27</sup>G. E. Crooks, "Path-ensemble averages in systems driven far from equilibrium," *Phys. Rev. E* **61**, 2361 (2000).
- <sup>28</sup>R. M. Neal, "Annealed importance sampling," *Stat. Comput.* **11**, 125 (2001).
- <sup>29</sup>S. Asakura and F. Oosawa, "On interaction between two bodies immersed in a solution of macromolecules," *J. Chem. Phys.* **22**, 1255 (1954).
- <sup>30</sup>K. Binder, P. Virnau, and A. Statt, "Perspective: The Asakura Oosawa model: A colloid prototype for bulk and interfacial phase behavior," *J. Chem. Phys.* **141**, 140901 (2014).
- <sup>31</sup>W. C. K. Poon, "The physics of a model colloid-polymer mixture," *J. Phys.: Condens. Matter* **14**, R859 (2002).
- <sup>32</sup>M. Dijkstra, J. M. Brader, and R. Evans, "Phase behaviour and structure of model colloid-polymer mixtures," *J. Phys.: Condens. Matter* **11**, 10079 (1999).
- <sup>33</sup>F. Lo Verso, R. L. C. Vink, D. Pini, and L. Reatto, "Critical behavior in colloid-polymer mixtures: Theory and simulation," *Phys. Rev. E* **73**, 061407 (2006).
- <sup>34</sup>D. J. Ashton and N. B. Wilding, "Three-body interactions in complex fluids: Virial coefficients from simulation finite-size effects," *J. Chem. Phys.* **140**, 244118 (2014).
- <sup>35</sup>D. J. Ashton, N. B. Wilding, R. Roth, and R. Evans, "Depletion potentials in highly size-asymmetric binary hard-sphere mixtures: Comparison of simulation results with theory," *Phys. Rev. E* **84**, 061136 (2011).
- <sup>36</sup>M. Dijkstra and R. van Roij, "Entropic wetting and many-body induced layering in a model colloid-polymer mixture," *Phys. Rev. Lett.* **89**, 208303 (2002).
- <sup>37</sup>M. Dijkstra, R. van Roij, R. Roth, and A. Fortini, "Effect of many-body interactions on the bulk and interfacial phase behavior of a model colloid-polymer mixture," *Phys. Rev. E* **73**, 041404 (2006).
- <sup>38</sup>T. Spyriouni, C. Tzoumanekas, D. Theodorou, F. Mueller-Plathe, and G. Milano, "Coarse-grained and reverse-mapped united-atom simulations of long-chain atactic polystyrene melts: Structure, thermodynamic properties, chain conformations, and entanglements," *Macromolecules* **40**, 3876 (2007).
- <sup>39</sup>A. Vrij, "Polymers at interfaces and the interactions in colloidal dispersions," *Pure Appl. Chem.* **48**, 471 (1976).
- <sup>40</sup>A. P. Gast, C. K. Hall, and W. B. Russel, "Polymer-induced phase separations in nonaqueous colloidal suspensions," *J. Colloid Interface Sci.* **96**, 251 (1983).
- <sup>41</sup>H. N. W. Lekkerkerker, W. C.-K. Poon, P. N. Pusey, A. Stroobants, and P. B. Warren, "Phase behaviour of colloid + polymer mixtures," *Europhys. Lett.* **20**, 559 (1992).
- <sup>42</sup>M. Schmidt, H. Löwen, J. M. Brader, and R. Evans, "Density functional theory for a model colloid polymer mixture: Bulk fluid phases," *J. Phys.: Condens. Matter* **14**, 9353 (2002).
- <sup>43</sup>J. M. Brader, R. Evans, and M. Schmidt, "Statistical mechanics of inhomogeneous model colloid-Polymer mixtures," *Mol. Phys.* **101**, 3349 (2003).
- <sup>44</sup>S. M. Oversteegen and R. Roth, "General methods for free-volume theory," *J. Chem. Phys.* **122**, 214502 (2005).
- <sup>45</sup>N. B. Wilding, "Critical-point and coexistence-curve properties of the Lennard-Jones fluid: A finite-size scaling study," *Phys. Rev. E* **52**, 602 (1995).
- <sup>46</sup>R. L. C. Vink and J. Horbach, "Grand canonical Monte Carlo simulation of a model colloid-polymer mixture: Coexistence line, critical behavior, and interfacial tension," *J. Chem. Phys.* **121**, 3253 (2004).
- <sup>47</sup>H. Oberhofer and C. Dellago, "Efficient extraction of free energy profiles from nonequilibrium experiments," *J. Comput. Chem.* **30**, 1726 (2009).
- <sup>48</sup>Y. Chen, "Another look at rejection sampling through importance sampling," *Stat. Probab. Lett.* **72**, 277 (2005).
- <sup>49</sup>S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso, and A. M. Stuart, "Importance sampling: Intrinsic dimension and computational cost," *Stat. Sci.* **32**, 405 (2017).
- <sup>50</sup>J. Liu, N. B. Wilding, and E. Luijten, "Simulation of phase transitions in highly asymmetric fluid mixtures," *Phys. Rev. Lett.* **97**, 115705 (2006).
- <sup>51</sup>D. J. Ashton and N. B. Wilding, "Grand canonical simulation of phase behaviour in highly size-asymmetrical binary fluids," *Mol. Phys.* **109**, 999 (2011).
- <sup>52</sup>M. Dijkstra, R. van Roij, and R. Evans, "Phase behavior and structure of binary hard-sphere mixtures," *Phys. Rev. Lett.* **81**, 2268 (1998).
- <sup>53</sup>M. Dijkstra, R. van Roij, and R. Evans, "Phase diagram of highly asymmetric binary hard-sphere mixtures," *Phys. Rev. E* **59**, 5744 (1999).
- <sup>54</sup>G. Odriozola, F. Jimenez-Angeles, and M. Lozada-Cassou, "Entropy driven key-lock assembly," *J. Chem. Phys.* **129**, 111101 (2008).
- <sup>55</sup>G. van Anders, D. Klotsa, N. K. Ahmed, M. Engel, and S. C. Glotzer, "Understanding shape entropy through local dense packing," *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4812 (2014).
- <sup>56</sup>C. Law, D. J. Ashton, N. B. Wilding, and R. L. Jack, "Coarse-grained depletion potentials for anisotropic colloids: Application to lock-and-key systems," *J. Chem. Phys.* **145**, 084907 (2016).
- <sup>57</sup>D. J. Ashton, R. L. Jack, and N. B. Wilding, "Porous liquid phases for indented colloids with depletion interactions," *Phys. Rev. Lett.* **114**, 237801 (2015).
- <sup>58</sup>A. Jusufi, M. Watzlawek, and H. Löwen, "Effective interaction between star polymers," *Macromolecules* **32**, 4470 (1999).
- <sup>59</sup>D. Marzi, C. N. Likos, and B. Capone, "Coarse graining of star-polymer-colloid nanocomposites," *J. Chem. Phys.* **137**, 014902 (2012).
- <sup>60</sup>D. A. Lenz, R. Blaak, C. N. Likos, and B. M. Mladek, "Microscopically resolved simulations prove the existence of soft cluster crystals," *Phys. Rev. Lett.* **109**, 228301 (2012).
- <sup>61</sup>N. B. Wilding and P. Sollich, "Demixing cascades in cluster crystals," *J. Chem. Phys.* **141**, 094903 (2014).
- <sup>62</sup>J. Largo and N. B. Wilding, "Influence of polydispersity on the critical parameters of an effective-potential model for asymmetric hard-sphere mixtures," *Phys. Rev. E* **73**, 036115 (2006).